R E P O R T R E S U M E S

ED 019 651 AL 000 970

COMPUTATIONAL SOCIOLINGUISTICS.

BY- SEDELOW, WALTER A., JR.
NORTH CAROLINA UNIV., CHAPEL HILL,INST.RES.SOC.SCI

PUB DATE NOV 67

EDRS PRICE MF-$0.25 HC-$0.32 6P.

DESCRIPTORS- *COMPUTATIONAL LINGUISTICS, *SOCIOLINGUISTICS, *LANGUAGE RESEARCH, BEHAVIORAL SCIENCE RESEARCH, COMPUTERS, *INTERACTION PROCESS ANALYSIS, MODELS,

THE USE OF THE COMPUTER MAY BE ONE OF THE WAYS IN WHICH VARIED LINGUISTIC INTERESTS (SOCIOLINGUISTICS, PSYCHOLINGUISTICS) COME TO BE RENDERED INTERRELATED AND EVEN INTELLECTUALLY COHERENT. (THE CRITERION OF COHERENCE IS SET HERE AT MONISM AS TO MODELS.) ONE OF THE AUTHOR'S MAJOR INTERESTS IS A SYSTEMATIC APPROACH TO SCIENTIFIC CREATIVITY, PARTICULARLY IN THE WORK OF SOCIAL-BEHAVIORAL SCIENTISTS. A PRECONDITION FOR FURTHER RESEARCH IS EXACTNESS OF DESCRIPTION OF LONG LANGUAGE STRINGS, SO THAT THEY MAY BE DISCRIMINATED ONE FROM ANOTHER WITH HIGH RELIABILITY AND SO THAT EVEN SLIGHT SHIFTS IN DISTRIBUTIONS OF CHARACTERISTICS MAY BE READILY DISCERNED. TO CHECK THE EMPIRICAL ADEQUACY OF HUMAN PERCEPTIONS AS TO CHARACTERISTICS OF LONG STRINGS OF SOCIAL ANALYSIS PROSE OVER AGAINST THE COMPUTATION RIGOR OF THE PROGRAMED MACHINE, AS WELL AS TO PROVIDE A SIGNIFICANT SET OF HUMAN DECISIONS FOR A MAN-MACHINE SYSTEM TO APPROXIMATE, THE AUTHOR IS ANALYZING SOME SUBSTANTIAL SEGMENTS OF "HUME'S HISTORY OF ENGLAND," A MAJOR WORK OF SOCIAL ANALYSIS. THE COMPARISONS WITH HUMAN JUDGMENTS WILL MAKE USE OF SOME OF THE AUTHOR'S EARLIER STUDIES OF THE SAME TEXT WITHOUT THE COMPUTER'S AID. THIS CASE STUDY SHOULD BE USEFUL, THE AUTHOR FEELS, IN FACILITATING THE SEMI-AUTOMATING OF A PROCEDURE FOR RATHER ROUTINELY CONDUCTING SUCH VERBAL ARTIFACT ANALYSIS. THE CATEGORIES EMPLOYED MAY BE FITTED INTO A MORE COMPREHENSIVE, SYSTEMATIC MODEL OF CREATIVITY AS IT OCCURS IN THE COMMUNICATION PROCESSES OF SOCIAL SCIENTISTS. A PROGRAMED MODELING OF THIS SORT NECESSARILY IMPLIES AN ATTENTION TO THE PRAGMATIC AND SEMANTIC ASPECTS OF ACTS OF COMMUNICATION. ACCORDINGLY, THIS IS ONE PHASE OF A COMPUTATIONAL PROJECT IN SOCIOLINGUISTICS EXEMPLIFYING A KIND OF APPROACH THAT CAN BE USED TO FIT TOGETHER LINGUISTIC AND SOCIOLOGICAL VARIABLES FOR THE SYSTEMATIC STUDY OF HUMAN INTERACTION. THIS PAPER APPEARED IN "RESEARCH PREVIEWS, VOLUME 14, NUMBER 2," NOVEMBER 1967, PUBLISHED BY THE INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE, THE UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NORTH CAROLINA 27514. (AUTHOR/AMM)

ED019651

RESEARCH PREVIEWS

Volume 14, No. 2

November 1967

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

"Computational Sociolinguistics"

by

Walter A. Sedelow, Jr.

"PERMISSION TO REPRODUCE THIS [redacted] MATERIAL HAS BEEN GRANTED BY WALTER A SEDELOW JR.
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE [redacted] OWNER."

Institute for Research in Social Science

The University of North Carolina

Chapel Hill

AL 000 970

COMPUTATIONAL SOCIOLINGUISTICS*

By way of background for a brief discussion of a computer-aided research effort in the sociology of language, it might be helpful to provide some orienting information about the comparatively new field of computational sociolinguistics.

Although there is now in process a substantial reaction against the unusually heavy emphasis in American linguistics on exclusively intersymbolic analysis, for many decades linguistics in this country tended to produce analyses which were primarily "syntactic" or intralingual, with comparatively little regard for the antecedent or consequent aspects of language behavior. There was however, understandably, relatively little concern for "semantics" and far less for "pragmatics," if one puts the matter in the widely employed Morrisean categories.

Under those circumstances there did develop, in some isolation from linguistics in this more confined sense, a number of subspecialties at least ancillary to linguistics; but they were located in other academic disciplines. Anthropological linguistics, to be sure, has been far closer to linguistics in its most straitened sense than have psycholinguistics and sociolinguistics; indeed, in some universities linguistics became almost as much the preserve of anthropology departments as earlier it had been that of the historical philologians in departments of languages and literatures. In America psycholinguistics has flourished for going on two decades, but it has done so rather independently of the linguistic "establishment." To date sociolinguistics has not had the intensity of fostering attention that the psychology of language has recently received, but to a considerable extent it does share with psycholinguistics the characteristic of a growth along its own lines rather than as an appendage of linguistics. In studying the social functions of language sociolinguists characteristically have made the boundaries of their analyses social rather than lingual, thus placing language events in a social setting rather than explicating social actions in models of linguistic contexts.

The application of large digital computers and programs to the analysis of data is today a characteristic of much linguistic research, both basic and applied, and irrespective of the departmental and disciplinary affiliations of the investigators. But taken together with revisions of concepts of linguistics, and the general growth of interest in systems approaches, the use of the computer may be one of the ways in which varied linguistic interests come to be rendered interrelated and, even, intellectually coherent. Here the criterion of coherence might be set at monism as to models.

---

*Dr. Sedelow is Dean, School of Library Science; Associate Professor, Department of Sociology and Department of Information Science; and Research Associate, Institute for Research in Social Science, University of North Carolina at Chapel Hill.

If, then, we define sociolinguistics to be the study of language-as-behavior in social context, computational sociolinguistics is the result of the application to that study of the power of the computer. As computational linguistics shares with traditional linguistics a conceptual isolating of language as an independent order of reality, so computational sociolinguistics may be said to share with the sociology of language the more comprehensive models of language in relation to other, circumambient social variables. At the various linguistics meetings held at the University of California at Los Angeles during the summer of 1966 a number of participants, and not least those from abroad, bespoke the importance of re-establishing in social context the analysis of language. Insofar as this mode of procedure comes to prevail, the sociology of language should find itself, increasingly, toward the center of linguistic research. Since the rigor and scientific character sought by many of the younger prominent linguists imply a multiplicity of parameters and a volume of data which, practically speaking, can by adequately coped with only be very large computational facilities, it may be urged that for linguistics in its more restricted senses the structuring for computational handling of research problems of a sociolinguistic character prospectively is also of considerable importance.

Although it would be impractical here to attempt to summarize the results or significance for general sociology of work to date in the sociology of language, much less its utility for linguistics, there is one possibly massive implication of computational sociolinguistics for the future of social behavior studies that might be said to demand mention. As we more intensively study social infrastructures perhaps the biggest hole in our analytic net is the absence of a tight examination of the verbal component of face-to-face interactions as part of a total kinesis of small groups. And without denying for a moment the feasibility, even the high desirability, of simultaneously employing models of social phenomena varying greatly in scale, it might be argued that a sine qua non for a science of human behavior is a comprehensive modeling of behavior within that social cell, the small group. For reliable, precise prediction of interactive behavior it seems self-evident that we need to cope analytically with the verbal elements in that behavior which constitute such a substantial part of it. For some features of the total interaction the Homunculus programs of the Gullahorns provide a useful approach, and for certain types of language exchange the General Inquirer of Philip Stone and others, is useful. But the General Inquirer requires precoding into fixed categories the word components of the strings of language to be studied. In addition to any questions we may have about the satisfactoriness of those categories, it is evident that given the amounts of verbal data to be processed we need a largely automated, reactive capability for that processing. A marriage of language-analytic procedures--computer based--with other highly automated procedures for characterizing human interaction should produce offspring that could grow to prodigious power in predicting behavioral detail even in nonexperimentally-structured settings. Just as it may seem narrow and unrealistic for linguists to have divorced studies of language from studies of the social conditions and correlates of its use, so we may assume it will come to seem narrow and unrealistic for social

scientists to divorce studies of human behavior from the linguistic models and techniques which pre-eminently apply to so much of that behavior *in vivo*.

* * * * *

One of my major interests is a systematic approach to scientific creativity, and, more particularly, to creativity in the work of social/behavioral scientists. The processes whereby new verbal models, or protomodels, emerge into the discourse of social analysis are an intermediate subject of attention in this effort. A more proximate target, and a precondition for further progress along certain lines, is exactness of description of long language strings, so that they may be discriminated one from another with high reliability and so that even slight shifts in distributions of characteristics may be readily discerned.

One of the criteria used in evaluating "artificial intelligence" research is the closeness of approximation to human decisions of its outputs. Since the quantities of verbal data to be studied are so large, it is important to automate the needed procedures as fully as possible, thus implying artificial intelligence rather than just the more obvious forms of verbal data processing. To check the empirical adequacy of human perceptions as to characteristics of long strings of social analysis prose over against the computation rigor of the programmed machine, as well as to provide a significant set of human decisions for a man-machine system to approximate, I am doing some analyses of substantial segments of Hume's *History of England* with the assistance of a set of programs generated in the Linguistics Analysis Project (earlier, the Stylistics Analysis Project), sponsored by the Office of Naval Research. The comparisons with human judgments will make use of some earlier studies of Hume's *History of England* that I undertook without the computer's aid.

Some of the dimensions that I have used in attempting to describe characteristics of Hume's lexicon, rhetoric, and mythology are also of high pragmatic significance, in the sense that they also may be used in measuring a personality and social/cultural variables that have impinged on its formation.

So it might be said that in this research the computations carried out with the aid of the computer should have relevance for characterizing Hume's *History of England*, a major work of social analysis; that this case study should be useful in facilitating the semi-automating of a procedure for rather routinely conducting such verbal artifact analysis; and that the categories employed may be fitted into a more comprehensive, systematic model of creativity as it occurs in the communication processes of social scientists. A programmed modeling of this sort necessarily implies an attention to the pragmatic and semantic aspects of acts of communication, and thus we have here one phase of a computational project in sociolinguistics exemplifying a kind of approach that can be used to fit together linguistic and sociological variables for the systematic study of human interaction.

Abstract of COMPUTATIONAL SOCIOLINGUISTICS

by Walter A. Sedelow, Jr.

The use of the computer may be one of the ways in which varied linguistic interests come to be rendered interrelated and, even, intellectually coherent. Here the criterion of coherence might be set at monism as to models. We define sociolinguistics to be the study of language-as-behavior in social context, and computational sociolinguistics as the result of the application to that study of the power of the computer. As we more intensively study social infrastructures, perhaps the biggest hole in our analytic net is the absence of a tight examination of the verbal component of face-to-face interactions as part of a total kinesis of small groups. For reliable, precise prediction of interactive behavior it seems self-evident that we need to cope analytically with the verbal elements in that behavior which constitute such a substantial part of it.

One of my major interests is a systematic approach to scientific creativity, and, more particularly, to creativity in the work of social/behavioral scientists. The processes whereby new verbal models, or protomodels, emerge into the discourse of social analysis are an intermediate subject of attention in this effort. A more proximate target, and a precondition for further progress along certain lines, is exactness of description of long language strings, so that they may be discriminated one from another with high reliability and so that even slight shifts in distributions of characteristics may be readily discerned. To check the empirical adequacy of human perceptions as to characteristics of long strings of social analysis prose over against the computation rigor of the programmed machine, as well as to provide a significant set of human decisions for a man-machine system to approximate, I am doing

some analyses of substantial segments of Hume's History of England primarily with the assistance of a set of programs generated in the Linguistics Analysis Project. The comparisons with human judgments will make use of some earlier studies of Hume's History of England that I undertook without the computer's aid. Some of the dimensions that I have used in attempting to describe characteristics of Hume's lexicon, rhetoric, and mythology are also of high pragmatic significance, in the sense that they also may be used in measuring a personality and social/cultural variables that have impinged on its formation.

So it might be said that in this research the computations carried out with the aid of the computer should have relevance for characterizing Hume's History of England, a major work of social analysis; that this case study should be useful in facilitating the semi-automating of a procedure for rather routinely conducting such verbal artifact analysis; and that the categories employed may be fitted into a more comprehensive, systematic model of creativity as it occurs in the communication processes of social scientists. A programmed modeling of this sort necessarily implies an attention to the pragmatic and semantic aspects of acts of communication, and thus we have here one phase of a computational project in sociolinguistics exemplifying a kind of approach that can be used to fit together linguistic and sociological variables for the systematic study of human interaction.